

Forecasting graduate student enrollment in university using regression analysis

Anggraini Dwiansyah¹, Imam Fahrurrozi¹, Muhammad Fakhurrifqi¹, Umar Farooq², Ganjar Alfian¹

¹Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta, Indonesia

²Faculty of Business and Law, Coventry University, Coventry, United Kingdom

Article Info

Article history:

Received Dec 15, 2024

Revised Feb 5, 2026

Accepted Feb 22, 2026

Keywords:

Accreditation

Linear regression

Machine learning

Monitoring

Prediction

ABSTRACT

The government ensures educational quality in universities through a quality assurance (QA) system implemented via accreditation, which evaluates both study programs and institutions. A key concern in accreditation is the decline in new student enrollment, making accurate predictions of enrollment numbers essential for quality assessment. This study proposes a linear regression (LR) model to forecast future university student enrollments based on enrollment figures from the previous year as input feature. Using a dataset from one of Indonesia's leading university spanning 2013 to 2023, the experimental results demonstrate that the LR model outperforms other regression techniques, including multi-layer perceptron (MLP), K-nearest neighbors (KNN), decision tree (DT), and random forest (RF). The LR model achieves R^2 values between 0.87 and 0.95, reflecting a strong linear relationship between current and future student numbers. It also delivers high accuracy, with root mean square error (RMSE) values ranging from 11.72 to 41.21 per year. The trained LR model has been integrated into a web-based system, offering data visualization and enrollment predictions to support university management in monitoring quality, addressing enrollment challenges, and facilitating informed decision-making.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ganjar Alfian

Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada

Yogyakarta 55281, Indonesia

Email: ganjar.alfian@ugm.ac.id

1. INTRODUCTION

Ensuring the quality of higher education is crucial for producing graduates capable of developing their potential and contributing valuable knowledge and technology to society. Governments implement quality assurance (QA) systems, often through accreditation processes, to uphold educational standards. In world-leading universities, these processes ensure that academic programs meet established criteria of excellence, facilitating global recognition, trust in academic credentials, and continuous institutional improvement. Accreditation is not a static process; rather, it involves recurring evaluations that challenge institutions to sustain and elevate their quality. Accreditation and QA frameworks serve as the backbone for higher education institutions, offering a structured way to measure and maintain excellence [1]. To this end, governments often establish national QA frameworks to uphold educational standards [2].

While specific metrics may vary globally, international accreditation frameworks fundamentally assess an institution's overall health and sustainability [1]-[3]. A critical component of this is financial viability, which is directly linked to the institution's ability to manage its student numbers. As noted by [4], university financial officers rely on accurate projections of new student enrollment to develop sound financial

plans. Therefore, monitoring and forecasting student enrollment is an essential activity for institutions to demonstrate the stability and long-term viability required to maintain their accreditation status.

Numerous studies have applied various forecasting methods to predict new student enrollment, employing different machine learning models and datasets. Various machine learning methods, such as single exponential smoothing, double exponential smoothing, random forest (RF), least squares, multi-layer perceptron (MLP), and LSTM, have been successfully applied in previous studies to predict new student enrollments, achieving different levels of accuracy [5]-[9]. In addition, more recent research has employed advanced models like WOASVR [10], ARIMA and DeepEnrollNet [11], Prophet [12], decision trees (DT) and neural networks [13], and classifiers such as RepTree [14], which have shown potential in capturing enrollment trends for better policy and resource planning. Complementing these studies, previous study [15] compares linear regression (LR) with Lasso, Elastic Net, SVR, and K-nearest neighbor (KNN) for predicting higher education KPIs, showing that LR achieved the best performance with 40% training and 60% testing split, as evaluated by mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), and coefficient of determination (R^2) metrics. Similarly, study by [16] utilized regression analysis to forecast student enrollment in private universities, showing that promotional costs significantly influence future admissions with a MAPE accuracy of 2.229%.

Based on previous studies, machine learning techniques, particularly LR, are effective for predicting future student numbers, with varying levels of accuracy depending on the dataset and method used [17]-[19]. Previous studies have not explored the application of LR for predicting future student enrollment numbers and compared its performance with other regression methods. Furthermore, the integration of such predictive models into a web-based application remains unexplored. Addressing this gap, this study develops a predictive model using LR to assist university management in maintaining enrollment numbers in alignment with accreditation standards. A key novelty of this research is the implementation of the model into a web-based system, enabling users to seamlessly monitor and track enrollment predictions. The system offers visual analytics, real-time monitoring, and data-driven insights to support decision-making processes and ensure the sustained quality of academic programs.

2. METHOD

The analysis process in this study consists of several key phases, including data collection, data preprocessing, model selection, model evaluation, and deployment [20]. As illustrated in Figure 1, this study began with data collection, focusing on new student enrollment data from postgraduate programs at a leading university in Indonesia over the past 11 years. The data processing phase followed, involving data cleaning procedures such as outlier detection and handling of missing values. In this study, LR and other machine learning models are utilized to forecast future university student enrollments using enrollment data from the preceding year. To structure the time series data effectively, a sliding window approach was employed to generate the input and output for the prediction models. The next step involved evaluating the algorithm's performance to identify the prediction model with the highest accuracy. Finally, the optimal prediction model was integrated into a web-based system designed to forecast new student enrollment numbers.

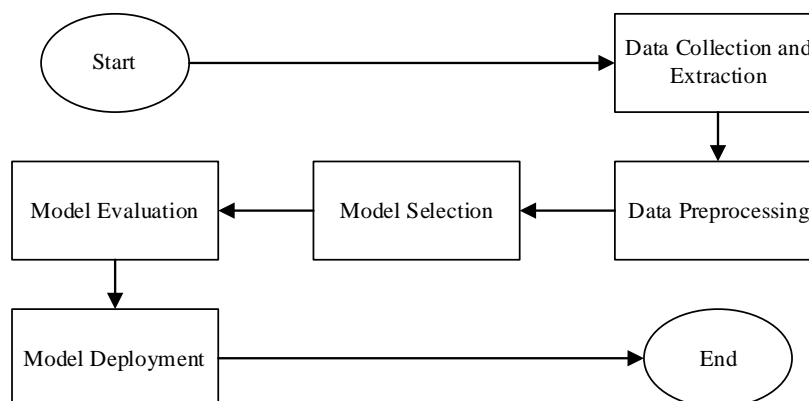


Figure 1. Flowchart of proposed study

2.1. Dataset

This study utilizes a time-series dataset comprising historical records of new student enrollments across 148 master's and doctoral programs at Universitas Gadjah Mada (UGM), Indonesia. Sourced from the UGM Directorate of Education and Teaching, the dataset includes features such as study program ID, name, level, faculty, and annual enrollment numbers spanning 11 years (2013–2023). With a total of 1,628 data points, the dataset provides yearly enrollment figures for each study program. Several supervised learning algorithms, including MLP, KNN, DT, and RF, will be applied to evaluate the data [21]. Additionally, the accreditation dataset is utilized to enhance the monitoring features by incorporating columns such as institution, accreditation rank, start date, and end date of the accreditation validity period.

2.2. Pre-processing stage

2.2.1. Sliding window method

Sliding window is a technique of processing time series or sequential data into smaller windows. Each window contains as much historical data as the predefined window size to be used as input that will predict the value of the first data of the next window [22]. Furthermore, the window will shift one step forward and continue until the time-series data has been divided into segments [23]. In this study, the sliding window technique plays a crucial role in preparing the dataset for time series forecasting. This method restructures the time series data into input (X) and output (Y) pairs, enabling supervised learning algorithms to train and generate predictions. By utilizing historical data within each window, the model learns patterns from previous years (X) to predict next year enrollment numbers (Y). This approach is fundamental for forecasting the number of new students in upcoming years based on historical trends. An illustration of the use of sliding window on a historical dataset of the number of new students with a window size of one is shown in Figure 2.

Data										
2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
53	38	37	59	44	49	57	61	62	79	64

Window Slides		
	X1	Y
2013	53	38
2014	38	37
2015	37	59
2016	59	44
2017	44	49
2018	49	57
2019	57	61
2020	61	62
2021	62	79
2022	79	64

Window Size = 1

Figure 2. Sliding window implementation on a timeseries dataset

2.3. Model selection

After undergoing data preprocessing, the datasets are utilized for model development by applying various algorithms and tuning parameters to identify the optimal model. In this study, several supervised machine learning algorithms are compared to determine the most effective approach for predicting new student enrollment numbers. The algorithms tested include LR, MLP, KNN, DT, and RF. Each algorithm is trained on the same training dataset and evaluated on the same test dataset to ensure a fair comparison and identify the best-performing prediction model.

2.3.1. Linear regression algorithm

LR is a machine learning-based method used to find a linear relationship between one or more predictors [24]. LR is part of supervised learning, which is used to predict a dependent variable (also known as the target or output) based on one or more independent variables (predictors or inputs). In its simplest form, LR assumes a linear relationship between the input (X) and output (Y), where the goal is to find the best-fitting line that minimizes the error between the predicted values and the actual values in the dataset [25]. The equation for simple LR can be represented in (1):

$$Y_i = a + bX \quad (1)$$

The objective of LR is to find the values of a (intercept) and b (slope) that minimize the difference between the actual data and the predicted values [26]. In this study, a simple LR model is used to predict the number of new students. For instance, the training data consists of a historical dataset where the current number of new students serves as the input (X), and the number of new students for the following year is the output (Y).

2.4. Model evaluation

Model evaluation is a stage conducted after the model is trained with training data to assess the performance of the machine learning model [27]. Model evaluation is done by utilizing test data to compare the prediction results with the actual output (Y) value of the test data. Evaluation metrics used in regression models include RMSE and R^2 . These metrics are used to measure how well the model represents the relationship between input and output variables. A smaller RMSE value indicates a better prediction model because it has a low error rate so that the model prediction is close to the actual value [28]. Meanwhile, the R^2 value, which is the square of the pearson correlation coefficient (r), ranges from 0 to 1. An R^2 value close to one indicates a better prediction model because it shows a positive relationship between variables.

2.4.1. Root mean square error

RMSE is an evaluation metric used to measure how well the regression model predicts the data by calculating the squared of the average prediction error. The smaller the RMSE value proves that the better the model because the prediction error is smaller [29]. RMSE is calculated using (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Predicted_i - Actual_i)^2}{n}} \quad (2)$$

2.4.2. Coefficient of determination

Another evaluation metric used in regression models is the R^2 . R^2 can be interpreted as the proportion of variance in the dependent variable that can be predicted from the independent variables with the best value being 1 and the worst value being 0. The calculation of R^2 is described by (3) to (4) which will produce the best value is 1 and the worst value is 0 [28].

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (Y - Y_i)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y_i)^2} \quad (4)$$

2.5. Model deployment

After the model evaluation shows good results, the selected model is then deployed. Model deployment is a stage in machine learning where models that have been trained and tested are prepared for use in a production environment so that they can be accessed and used by end-users. At this stage, the model is no longer just in development or experimentation, but can be used to predict the number of new students by implementing the model into a web-based system. System development utilizes streamlit library with the python programming language. The system is built with the aim of receiving input from end-users and producing output in the form of predicting the number of new students that can be seen by end-users. Figure 3 shows the architecture of the web-based system.

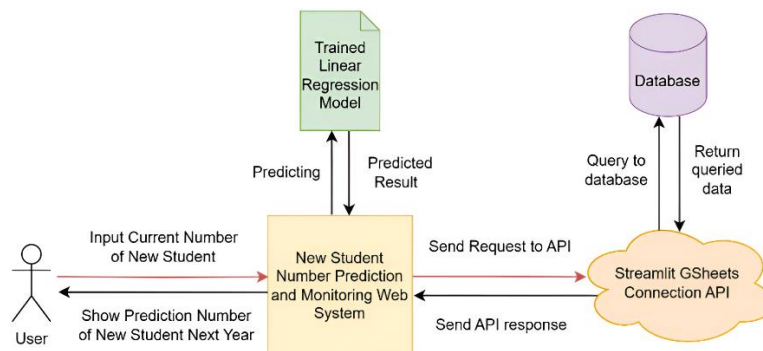


Figure 3. Architecture of web-based system

3. RESULTS AND DISCUSSION

This section presents the results of the prediction model developed using the LR method to predict new student enrollment in the postgraduate programs at Universitas Gadjah Mada. The model was evaluated using two primary metrics, namely, RMSE and R^2 , which helped assess the accuracy and the linear relationship between the current and future student enrollment figures. In addition, the web-based system will display data visualization of prediction and monitoring results using a LR model.

3.1. Performance comparison of regression methods

Table 1 presents the proposed LR model for predicting future student enrollment, compared with MLP, KNN, DT, and RF. These models were selected based on the dataset's characteristics and the objective of achieving accurate predictions for the following year's student numbers. The model performance was evaluated using two metrics: R^2 and RMSE, with the same test data applied to determine the best-performing model. The LR model outperformed the others, achieving the lowest RMSE of 12.006 and the highest R^2 of 0.943.

Although Table 1 shows the proposed LR model achieving the best overall performance—with the lowest RMSE of 12.006 and highest R^2 of 0.943—this finding appears to contrast with the yearly performance trends presented in Table 2. While the LR model demonstrates relatively consistent performance over time, models such as DT and RF outperformed LR in several individual years—particularly between 2014 and 2019—by achieving lower RMSE values and higher R^2 scores. This inconsistency suggests that Table 1 reflects average or cumulative performance across the entire dataset, favoring models with stable outputs, whereas Table 2 provides a more detailed view of performance variability from year to year. Consequently, while LR offers a simple and interpretable solution with strong overall accuracy, other models may be more suitable in specific contexts or time periods. Future research could explore hybrid modeling approaches that combine multiple algorithms to better handle variations in enrollment trends. Additionally, improvements may include integrating more relevant variables or utilizing more sophisticated regression techniques to more effectively capture unpredictable shifts in enrollment behavior [30].

Table 1. Model performance comparison results

Model	RMSE	R^2
MLP	12.161	0.943
KNN	15.287	0.943
DT	19.948	0.843
RF	17.688	0.876
Proposed LR	12.006	0.943

Table 2. Evaluation of regression models from 2013 to 2022

Current year	LR		MLP		KNN		DT		RF	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
2013	41.21	0.5	41.04	0.51	39.87	0.54	37.06	0.6	37.5	0.59
2014	14.12	0.92	13.94	0.92	12.18	0.94	6.04	0.98	7.11	0.98
2015	14.62	0.91	14.87	0.9	14.84	0.9	7.04	0.98	9.51	0.96
2016	17.98	0.87	17.98	0.87	18.15	0.86	5.33	0.99	9.39	0.96
2017	12.26	0.93	12.23	0.93	18.07	0.85	8.31	0.97	9.66	0.96
2018	11.72	0.95	11.77	0.95	20.7	0.84	6.94	0.98	9.07	0.97
2019	12.02	0.94	13.92	0.93	20.19	0.86	5.83	0.99	8.7	0.97
2020	15.72	0.89	15.46	0.89	22.71	0.77	9.25	0.96	11.33	0.94
2021	13.01	0.92	12.6	0.92	19.89	0.8	6.67	0.98	9.13	0.96
2022	15.3	0.9	14.92	0.9	21.02	0.81	7.99	0.97	10.77	0.95

The superior performance of the LR model (Table 1) in this study can be attributed to the simplicity of the input feature, which consists solely of the previous year's enrollment data. Given that the relationship between current and future enrollment tends to follow a strong linear trend, LR is well-suited to capture this pattern effectively. In contrast, more complex models such as MLP, KNN, DT, and RF are designed to model nonlinear relationships and may not perform optimally when applied to datasets with a single, linearly related predictor. As a result, the added complexity of these models does not yield significant performance benefits and may even reduce accuracy, as reflected in their higher RMSE values compared to LR.

3.2. Practical application

The evaluated model is then visualized through a web application built using the Streamlit library in Python, designed to predict the number of new students in the upcoming years and offer additional features for monitoring student enrollment. The system implementation includes both database and interface

development. The input data received is processed using the pre-trained prediction model, and the results are returned as response data to the web-based system, where they are displayed on the prediction page along with the monitoring features. This prediction page serves as the main feature of the system, showcasing the forecasted and monitored student enrollment numbers for future years. Users will be prompted to provide the necessary input data, after which the system processes it and presents the prediction results on the page.

3.2.1. Database implementation

In this study, the database is implemented using Google Sheets as a data storage platform with the help of the Streamlit Gsheets Connection Library. Google Sheets was chosen because this platform allows easy and flexible access, and supports direct collaboration between users and the system [31]. The data used for model training and prediction of the number of new students is stored in Google Sheets, where each row represents data for a particular year that includes the number of new students and other related variables. The system will automatically access and update the data in Google Sheets according to the prediction needs of the coming year.

3.2.2. User interface implementation

The user interface is built using Streamlit library, which allows the creation of interactive web applications quickly and easily [32]. The main page of the application serves to input the required data, such as the number of new students in previous years. After the data is entered, the system will process and display the prediction results automatically. In addition, the application is also equipped with a graph monitoring feature that shows the development of the number of new students from year to year, as well as predictions for future years. The prediction calculator page displays an interface to predict the number of new students in the next few years based on the current number of new students entered as shown in Figure 4. Figure 4 showed a declining trend in the predicted number of postgraduate enrollments over the years. This pattern might be due to several factors, such as the increasing number of postgraduate programs offered across various universities, which has intensified competition and dispersed prospective applicants. Additionally, many graduates are now more inclined to enter the workforce, especially in industry, rather than pursue further academic studies.

In addition, there is also a monitoring page for all study programs aimed at predicting the number of students in the next few years and monitoring the results of these predictions. This prediction uses historical data and calculations with certain formulas to produce predictions for monitoring the number of students to come as shown in Figure 5. The implementation of this user interface makes it easier for users to understand and interpret the prediction results and provides an interactive user experience.

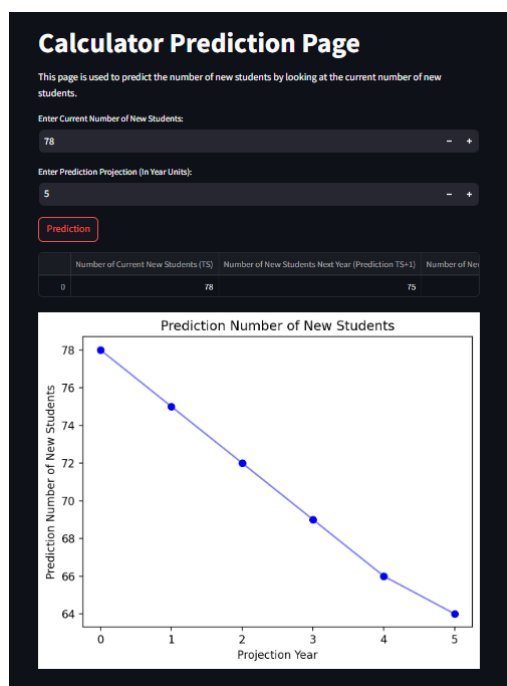


Figure 4. Calculator page

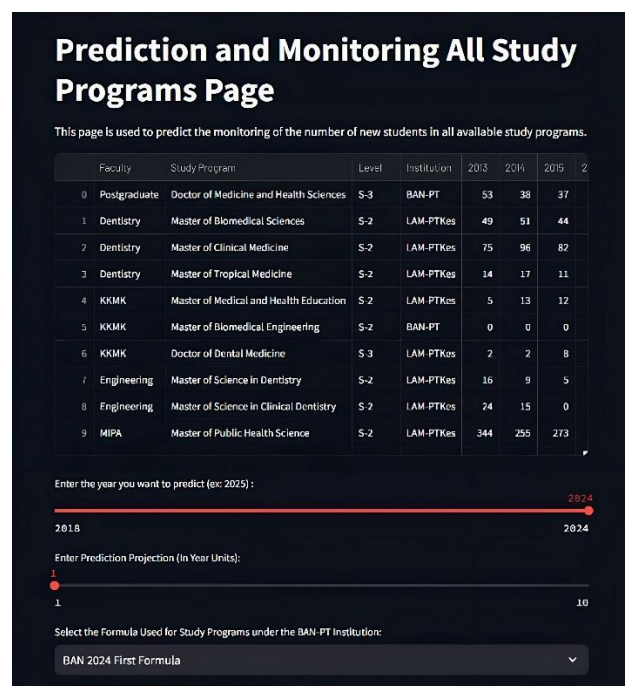


Figure 5. Prediction and monitoring page

4. CONCLUSION

This study has successfully developed a web-based system to predict and monitor new student enrollment using the LR method. The LR model was chosen after rigorous comparison with other regression techniques, including MLP, KNN, DT, and RF. This study uses enrollment data from the preceding year as the input feature to predict university enrollment for the following year. Experimental results demonstrated the superior performance of the LR model, which achieved the highest R² value of 0.943 and the lowest RMSE of 12.006, indicating its strong predictive capability for enrollment trends. The integration of the trained LR model into a user-friendly web application using Streamlit offers practical value beyond theoretical contribution. University administrators can now access an interactive platform that visualizes enrollment trends and generates forecasts, enabling data-driven strategies to address fluctuations in student intake. This aligns with the increasing importance of QA in higher education, where accurate and timely enrollment predictions are crucial for maintaining accreditation standards and ensuring institutional sustainability. The significance of this research lies in its contribution to educational management through the application of a simple yet effective predictive model, embedded in an accessible tool for decision support. By facilitating proactive responses to enrollment challenges, the system empowers universities to better plan academic resources, manage budgets, and uphold educational quality.

This study has several limitations, most notably the reliance on a single input feature—enrollment data from the previous year—which may reduce the model’s ability to generalize across varying conditions. Additionally, the use of basic regression techniques may not fully capture complex enrollment patterns influenced by multiple factors. Future enhancements may include updating and expanding the dataset, the use of diverse evaluation metrics, incorporating sophisticated predictive models for more complex trend analysis, and improving the system interface with additional features to monitor other relevant indicators, such as program accreditation and tuition fees. These enhancements will further improve the system's accuracy, usability, and ability to support decision-making.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anggraini Dwiansyah		✓	✓	✓	✓	✓		✓	✓	✓	✓			
Imam Fahrurrozi	✓				✓				✓			✓		✓
Muhammad Fakhurrifqi		✓				✓			✓		✓	✓		
Umar Farooq		✓		✓		✓	✓			✓		✓		✓
Ganjar Alfian	✓	✓				✓			✓	✓		✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [GA], upon reasonable request.




REFERENCES

- [1] H. Magd and S. K. Kunjumammed, *Global perspectives on quality assurance and accreditation in higher education institutions*. Advances in Higher Education and Professional Development. IGI Global, 2022, doi: 10.4018/978-1-7998-8085-1.




- [2] M. Kayyali, *Quality Assurance and Accreditation in Higher Education: Issues, Models, and Best Practices*. Cham: Springer Nature Switzerland, 2024, doi: 10.1007/978-3-031-66623-0.
- [3] S. Schwarz and D. F. Westerheijden, *Accreditation and Evaluation in the European Higher Education Area*, Springer Science & Business Media, vol. 5, 2004, doi: 10.1007/1-4020-2797-4.
- [4] D. Hossler and B. Bontrager, *Handbook of strategic enrollment management*. Jossey-Bass, 2015.
- [5] Q. Chen, "A Comparative Study on the Forecast Models of the Enrollment Proportion of General Education and Vocational Education," *International Education Studies*, vol. 15, no. 6, pp. 1–18, Nov. 2022, doi: 10.5539/ies.v15n6p109.
- [6] P. D. P. Silitonga, H. Himawan, and R. Damanik, "Forecasting acceptance of new students using double exponential smoothing method," *Journal of Critical Reviews*, vol. 7, no. 1, pp. 300–305, Jan. 2020, doi: 10.31838/jcr.07.01.57.
- [7] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [8] A. J. Ikuomola and A. A. Adebisi, "A comparison of fuzzy time series and least-square method in forecasting students' enrolment," *Journal of Natural Sciences Engineering and Technology*, vol. 10, no. 1, pp. 1–12, 2011, doi: 10.51406/jnset.v10i1.1280.
- [9] F. E. James and J. L. Weese, "Neural Network-Based Forecasting of Student Enrollment With Exponential Smoothing Baseline and Performance Analysis," in *2022 ASEE Annual Conference & Exposition*, Minneapolis, MN: ASEE Conferences, 2022, doi: 10.18260/1-2--41751.
- [10] S. Yang, H. C. Chen, W. C. Chen, and C. H. Yang, "Student Enrollment and Teacher Statistics Forecasting Based on Time-Series Analysis," *Computational Intelligence and Neuroscience*, pp. 1–15, Sep. 2020, doi: 10.1155/2020/1246920.
- [11] S. K. Sharma, "An instructional emperor pigeon optimization (IEPO) based DeepEnrollNet for university student enrolment prediction and retention recommendation," *Scientific Reports*, vol. 14, no. 1, pp. 1–24, Dec. 2024, doi: 10.1038/s41598-024-81181-9.
- [12] U. B. Patayon and R. V. Crisostomo, "Time Series Analysis on Enrolment Data: A case in a State University in Zamboanga del Norte, Philippines," in *2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Depok, Indonesia: IEEE, Oct. 2022, pp. 13–18, doi: 10.1109/ICACSIS56558.2022.9923436.
- [13] L. Yang, L. Feng, L. Zhang, and L. Tian, "Predicting freshmen enrollment based on machine learning," *Journal of Supercomputing*, vol. 77, no. 10, pp. 11853–11865, Oct. 2021, doi: 10.1007/s11227-021-03763-y.
- [14] B. Ujkani, D. Minkovska, and L. Stoyanova, "A Machine Learning Approach for Predicting Student Enrollment in the University," in *2021 XXX International Scientific Conference Electronics (ET)*, Sozopol, Bulgaria, Sep. 2021, pp. 1–4, doi: 10.1109/ET52713.2021.9579795.
- [15] A. Abdelhadi, S. Zainudin, and N. S. Sani, "A Regression Model to Predict Key Performance Indicators in Higher Education Enrollments," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 454–460, 2022, doi: 10.14569/IJACSA.2022.0130156.
- [16] S. P. Sipayung, Z. Situmorang, Z. Matondang, and M. Sagala, "Forecasting Analysis of New Student Candidate Admissions Using the Simple Linear Regression Method," *Jurnal Teknik Indonesia*, vol. 2, no. 02, pp. 64–69, Oct. 2023, doi: 10.58471/ju-ti.v2i02.561.
- [17] Y. Yamasari, N. Rochmawati, R. E. Putra, A. Qoiriah, Asmunin, and W. Yustanti, "Predicting the Students Performance using Regularization-based Linear Regression," in *2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Surabaya, Indonesia, Oct. 2021, pp. 1–5, doi: 10.1109/ICVEE54186.2021.9649704.
- [18] E. Aditya, Z. Situmorang, B. H. Hayadi, M. Zarlis, and Wanayumini, "New Student Prediction Using Algorithm Naive Bayes and Regression Analysis in Universitas Potensi Utama," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, Prapat, Indonesia: IEEE, Oct. 2022, pp. 1–6, doi: 10.1109/ICORIS56080.2022.10031391.
- [19] E. V. Maliberan, "Forecasting Enrolment Data of Surigao del Sur State University, Philippines using Regression Analysis and Multiplicative Decomposition Model," *International Journal of Advanced Science Computing and Engineering*, vol. 3, no. 1, pp. 1–9, May 2021, doi: 10.62527/ijasce.3.1.26.
- [20] J. Heit, J. Liu, and M. Shah, "An architecture for the deployment of statistical models for the big data era," in *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, Dec. 2016, pp. 1377–1384, doi: 10.1109/BigData.2016.7840745.
- [21] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons.B*, vol. 4, pp. 51–62, Dec. 2017, doi: 10.20544/horizons.b.04.1.17.p05.
- [22] I. Pisa, I. Santín, J. L. Vicario, A. Morell, and R. Vilanova, "Data preprocessing for ANN-based industrial time-series forecasting with imbalanced data," in *2019 27th European Signal Processing Conference (EUSIPCO)*, A Coruna, Spain, Sep. 2019, pp. 1–5, doi: 10.23919/EUSIPCO.2019.8902682.
- [23] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
- [24] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.
- [25] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, 5th Edition. in Wiley Series in Probability and Statistics. Hoboken, NJ, 2015. [Online]. Available: <https://sadbhavnapublications.org/research-enrichment-material/2-Statistical-Books/Regression-Analysis-by-Example.pdf>.
- [26] D. York, N. M. Evensen, M. L. Martinez, and J. De B. Delgado, "Unified equations for the slope, intercept, and standard errors of the best straight line," *American Journal of Physics*, vol. 72, no. 3, pp. 367–375, Mar. 2004, doi: 10.1119/1.1632486.
- [27] J. A. Pruneski *et al.*, "The development and deployment of machine learning models," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 30, no. 12, pp. 3917–3923, Dec. 2022, doi: 10.1007/s00167-022-07155-4.
- [28] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, pp. 1–24, Jul. 2021, doi: 10.7717/PEERJ-CS.623.
- [29] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.
- [30] G. A. Dimaculangan, "Using Regression Analysis to Forecast the Factors Affecting the Enrollment of a Tertiary School," *Asia Pacific Journal of Multidisciplinary Research*, vol. 2, no. 1, pp. 121–130, 2014.
- [31] M. Kosmala, M. Foley, T. R. Kurfess, and K. S. Saleeby, "Evaluation of No-Cost MTConnect-Enabled Machine Monitoring Architecture in Google Sheets," in *IET Conference Proceedings*, vol. 2023, no. 16, pp. 138–143, Oct. 2023, doi: 10.1049/icp.2023.1745.
- [32] M. Khorasani, M. Abdou, and J. H. Fernández, "Getting Started with Streamlit," in *Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework*, Berkeley, CA: Apress, 2022, pp. 1–30, doi: 10.1007/978-1-4842-8111-6_1.

BIOGRAPHIES OF AUTHORS






Anggraini Dwiansyah    obtained her Bachelor's degree in the Applied Undergraduate Program of Software Engineering Technology, Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Indonesia. Her research areas are data mining, machine learning, and web development. She can be contacted at email: anggraini.d@mail.ugm.ac.id.






Imam Fahrurrozi    received his Doctorate degree from the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. Additionally, he serves as a full-time lecturer at the Department of Electrical Engineering and Informatics in the Vocational College of Universitas Gadjah Mada, Indonesia. His research areas are artificial intelligence, IoT, machine learning, and recommender systems. He is a co-founder of KIRANA (Kios Rakyat Indonesia), a technology-based e-commerce company for small and medium enterprises (UMKM). He can be contacted at email: imam.fahrurrozi@ugm.ac.id.






Muhammad Fakhrrurifqi    currently works as Lecturer at Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Indonesia since 2012. He earned Master's degrees in Computer Science from Universitas Gadjah Mada and is currently pursuing a Ph.D. in the same field at the university. His research interests include artificial intelligence, information system, and health informatics. He can be contacted at email: rifqi_ilkom@mail.ugm.ac.id.



Umar Farooq    is currently working as a lecturer of Business Analytics at the School of Strategy and Leadership, College of Business and Law, Coventry University, United Kingdom. Previously, he worked as Assistant Professor for five years at Gulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan. He received his Masters and Ph.D. in Industrial and Systems Engineering from Dongguk University, South Korea in 2017. He has vast experience in the fields of supply chain management, operations management, and business analytics. His research interests are in RFID, IoT, food supply chains, and machine learning, and analytics. He can be contacted at email: ad9061@coventry.ac.uk.



Ganjar Alfian    has been serving as an Assistant Professor at the Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Indonesia, since 2022. Before this, he spent five years as an Assistant Professor at Dongguk University in Seoul, Republic of Korea. In July 2017, he was a short-term Visiting Researcher at VSB-Technical University of Ostrava in the Czech Republic. He received Dr.Eng, degrees from the Department of Industrial and Systems Engineering, Dongguk University, Seoul, South Korea, in 2016. His research interests include artificial intelligence, RFID, IoT, machine learning, carsharing, and simulation. He is listed in the World's Top 2% Scientist 2022, 2024 and 2025 by Stanford University. He can be contacted at email: ganjar.alfian@ugm.ac.id.